OXFORD

Sequence analysis

Linking sequence patterns and functionality of alpha-helical antimicrobial peptides

Igor E. Eliseev [iD] [1],*,[†], **Ivan N. Terterov**[1],[†], **Anna N. Yudenko**[1] and **Olga V. Shamova**[2]

[1]Nanobiotechnology Lab, St Petersburg Academic University, St Petersburg 194021, Russian Federation and [2]Department of General Pathology and Pathophysiology, Institute of Experimental Medicine, St Petersburg 197376, Russian Federation

*To whom correspondence should be addressed.

[†]The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

Associate Editor: John Hancock

Abstract

Motivation: The rational design of antimicrobial peptides (AMPs) with increased therapeutic potential requires deep understanding of the determinants of their activities. Inspired by the computational linguistic approach, we hypothesized that sequence patterns may encode the functional features of AMPs.

Results: We found that α-helical and β-sheet peptides have non-intersecting pattern sets and therefore constructed new sequence templates using only helical patterns. Designed peptides adopted an α-helical conformation upon binding to lipids, confirming that the method captures structural and biophysical properties. In the antimicrobial assay, 5 of 7 designed peptides exhibited activity against Gram(+) and Gram(−) bacteria, with most potent candidate comparable to best natural peptides. We thus conclude that sequence patterns comprise the structural and functional features of α-helical AMPs and guide their efficient design.

Contact: eliseev@spbau.ru

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Antimicrobial peptides (AMPs), found in nearly all multicellular organisms, are promising compounds for the development of new therapeutics against multidrug-resistant bacteria (Zasloff, 2002). Remarkably, AMPs do not induce a widespread bacterial resistance such as those that observed for conventional antibiotics. The unique properties of AMPs presumably originate from the millions of years of host and pathogen co-evolution which has shaped both the defense peptides repertoire and their mechanism of action (Peschel and Sahl, 2006). Unlike conventional antibiotics, cationic AMPs mostly target the inner membrane of bacterial cells. The exact mechanisms of membrane disruption vary among AMPs, but they all involve the formation of distinct amphipathic conformation on the lipid membrane (Powers and Hancock, 2003). This membrane-bound state is maintained by a secondary structure that naturally divides AMPs into three dominant structural classes: α-helical, β-sheet and extended peptides (Nguyen *et al.*, 2011).

Numerous efforts have been made to design new potent peptides with better therapeutic properties (Fjell *et al.*, 2011). One way is to enhance a natural AMP by altering its sequence either rationally (Chen *et al.*, 2005; Tossi *et al.*, 1997) or using high-throughput screening (Hilpert *et al.*, 2005), but these methods restrict the sequence space to the variants of the progenitor peptide. An alternative strategy employs artificial intelligence to guide *in silico* design, which proved to be effective in the *de novo* generation of highly active short peptides (Cherkasov *et al.*, 2009). Loose and colleagues proposed a more intuitive 'linguistic' approach, which explores an analogy between amino acid sequences and sentences in a language and derives certain grammar rules for AMPs, expressed as specific sequence patterns (Loose *et al.*, 2006). They synthesized peptides obeying these grammar rules and showed that some of them were active, proving that sequence patterns can recognize the determinants of antimicrobial activity.

The linguistic method does not distinguish structural classes. However, sequence patterns may encode certain structural features crucial for the mechanism of antibacterial action. We hypothesized that patterns derived from differently structured peptides may also be dissimilar, and their fusion in a new sequence can interfere with the formation of a functional conformation thus hindering the activity. Therefore, the rational design requires a procedure which would implement the properties carried by patterns and fully reproduce the structure and mechanism of natural antimicrobials. Here, we proposed an approach that uses patterns from the structurally homogeneous α-helical AMPs, preserves their topology and yields highly potent candidates.

2 Materials and methods

Sequences of AMPs exhibiting activity against Gram(+) and Gram(–) bacteria and having overall non-negative charge were selected from the AMP Database v. 2 (Wang *et al.*, 2009). The structural annotation was available for ∼24% of the peptides. We considered the two subsets of this group: α-helical peptides (273 sequences), comprising more than 70% of AMPs with known structure and *β*-sheet peptides (64 sequences). The sequences of AMPs and structure elucidation methods used in each case are summarized in Supplementary Material, Supplementary Tables S1 and S2.

We call *pattern* any string that begins and ends with a residue and contains an arbitrary combination of residues and *wildcards* ('.')—positions that may be occupied by any residue. Teiresias pattern discovery algorithm (Rigoutsos and Floratos, 1998) requires parameters L, W and K. L is the minimum number of residues in a pattern. The parameter W controls the density of wildcards so that W-L is the maximum number of wildcards in any sub-pattern that contains L residues. K denotes the minimal *support*, which is the number of times the given pattern appears in the database. The parameter settings were $L = 3$, $W = 10$ and $K = 3$. The larger values of W did not give a significant increase in the number of patterns while larger values of L decreased the number of patterns dramatically.

Details of the experimental methods, including peptide synthesis, structural analysis and antimicrobial assays, are given in Supplementary Material.

3 Results

3.1 Natural AMPs have characteristic sequence patterns
We used Teiresias algorithm to find sequence patterns in non-homologous AMP sequences (Loose *et al.*, 2006; Rigoutsos and Floratos, 1998). However, it is clear that patterns found in the database may include those that are not characteristic to AMPs but have high *support* as a coincidence. We estimated the minimal support for a pattern to consider it specific by comparing native sequences with *shuffled* variants generated by random permutations of residues. We applied Teiresias to the native and shuffled sets and focused on the number of patterns as a function of support. As seen from Figure 1, the number of patterns demonstrates quasi-exponential decay with the support and the slope of this dependence is different for the shuffled and native sets.

Thus, we can define a certain threshold for the minimal support of a specific pattern by computing mean maximal support in shuffled sets, which gives 30.5 ± 0.4 (SE, $n = 100$) for α-helical peptides. Patterns with support above the threshold are likely associated with an amphipathic structure and antibacterial activity.

3.2 Pattern analysis and validation
The 76 patterns with support ≥ 30 (given in Supplementary Material, Supplementary Table S3) were mapped to the sequences
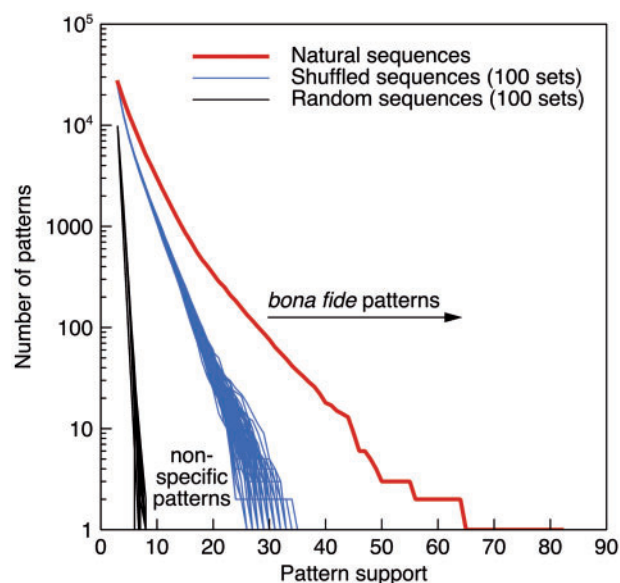
Fig. 1. Natural AMPs have characteristic sequence patterns. Number of patterns as a function of their support for α-helical AMPs. Analysis of shuffled and random sequences allows defining a certain threshold (support ∼30) separating non-specific and *bona fide* patterns

of α-helical AMPs from which they were derived. Number of patterns in each peptide and the distribution of pattern incidence among AMPs are given in Supplementary Table S1 and Supplementary Figure S1, respectively. On average, each natural peptide contained 7.9 patterns. However, 24% of peptides (66) did not contain any patterns with support ≥ 30. This typically occurred with members of certain families, e.g. *temporins*, *gramicidins* and *clavanins*. Although they obviously had well defined sequence motifs, such patterns were infrequently encountered outside the given family, which prevented them from gaining necessary support.

Some of the top-ranked patterns invite a simple structural interpretation. For instance, three most abundant patterns, K...K...K, K...K.K and K.K...K, encode a charged patch on one side of an α-helix. Similarly, patterns like A...A...A and A...A.A define a hydrophobic surface. Indeed, some of our patterns perfectly match sequence templates previously identified by calculating positional frequencies on helical wheel diagrams (Zelezetsky and Tossi, 2006). Graphical representation of patterns in several well-known AMPs may be found in Supplementary Figure S2.

We used bootstrapping to obtain a statistical estimate of variation in pattern supports. From the set of 273 α-helical peptides, we constructed 100 sets of the same size using random sampling with replacement. We then run Teiresias on these resampled sets with the same parameters as we used for the original dataset. The mean support and its SD for each pattern are given in Supplementary Table S3. Top patterns from bootstrapped sets matched those from the original set, and the coefficient of variation of support was ∼0.2, that indicates the reliability of specific patterns that we use.

3.3 α-helical and *β*-sheet AMPs have different patterns
Comparison of patterns from α-helical and *β*-sheet peptides revealed that only 0.2% of patterns are present in both subsets, and the number of common patterns rapidly decays with support. Particularly, for supports larger than 8, much less than the threshold, there are no patterns shared by α and β subsets. This finding is not surprising *per*

se, as it is well known that certain amino acids and sequence motifs determine secondary structure propensity for polypeptide chains (Finkelstein and Ptitsyn, 2002). However, returning to the linguistic metaphor, these results confirm our conjecture that differently structured AMPs represent different languages, and thus, their fusion in a synthetic peptide would scramble functional membrane-bound conformation.

3.4 Patterns facilitate the identification of novel AMPs

Having established that α-helical AMPs possess characteristic sequence patterns, we wondered whether these patterns can facilitate the identification of novel AMPs in databases not dedicated to antimicrobials. In a model computational experiment, we took all peptides from Swiss–Prot with length 10–30 a.a. and clustered them at 70% identity with Cd-hit program (Li and Godzik, 2006), which yielded 2454 peptides of different function and origin, not specifically enriched for AMPs. We then mapped specific patterns derived from helical AMPs (support ≥30) to the Swiss–Prot peptides and found that the average number of patterns per peptide was only 0.9 compared to 7.9 for AMPs (Supplementary Fig. S1). The 115 peptides containing >5 patterns were selected as potentially antimicrobial.

Strikingly, the analysis revealed that 65% (75) of these peptides had experimentally confirmed antibacterial activity, 15% (17) are presumably antibacterial based on homology and their identification in skin secretions and only 20% (23) were unrelated to AMPs. From the 75 active peptides, 18 belonged to the α-helical set used for pattern discovery, 49 sequences from APD3 (Wang *et al.*, 2016) were identified *de novo* and 8 peptides were not previously listed in APD3 (i.e. *grammistins*). Moreover, some of remaining 23 peptides can also be active. For example, one of them, *P13282*, is a fragment of histone H2B.2 from sea urchin sperm, appears an interesting candidate because of the potential role of histone fragments in host defense (Kawasaki and Iwamuro, 2008). The sequences, Uniprot and APD IDs and corresponding references are given in Supplementary Table S4.

3.5 A new strategy for AMP design

To design new synthetic peptides, we used patterns from α-helical peptides, the largest class of AMPs with annotated structure and simple topology without disulfide bonds. First, patterns with high support (≥30) were merged into templates of the fixed length 20, so that the last residue of the preceding pattern was the first residue of the following one as shown in Figure 2, giving 10^5 possible variants. Then for each template, we generated sequences by substituting the wildcards in patterns with random residues. The resulting set was filtered to keep peptides with net charge (+4e ± 2e) and number of hydrophobic residues (10 ± 4) close to values observed for natural α-helical peptides, and then clustered against the database of non-redundant protein sequences at 40% homology using Cd-hit (Li and Godzik, 2006). From 1270 sequences showing no homology with natural AMPs and other proteins we selected peptides with a high α-helical propensity predicted with Psipred (Jones, 1999) and synthesized seven candidates lacking cysteine, methionine and tryptophan (see Table 1 for sequences).

3.6 Designed peptides adopt an α-helical conformation in a lipid environment

Since the synthetic AMPs consisted of patterns derived from natural ones with an α-helical structure, we expected them to be α-helical as well. To test this hypothesis, we analyzed peptides by circular
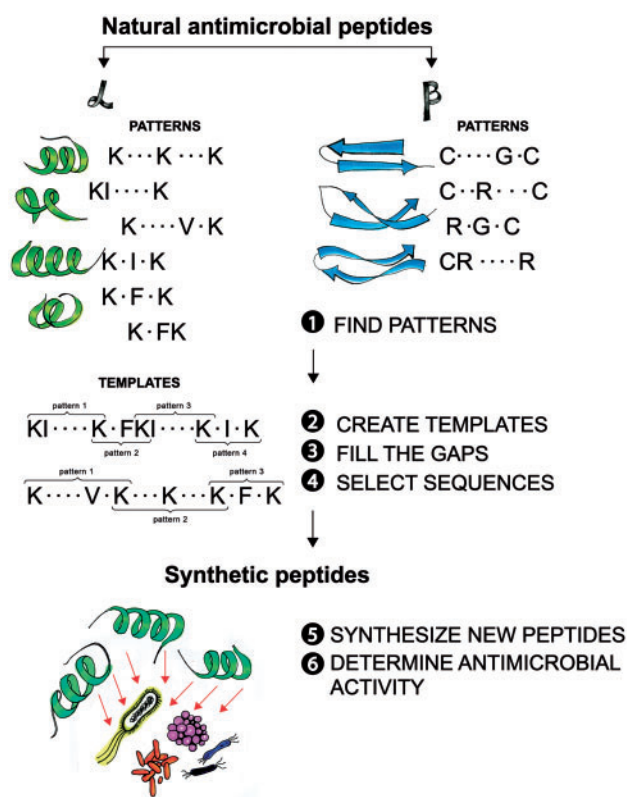
Fig. 2. A new strategy for AMP design. The sequence patterns derived from natural α-helical AMPs were joined to create templates. Then the wildcards were replaced by random residues, peptides were synthesized and assayed for antibacterial activity

dichroism spectroscopy in buffer solution alone and in membrane environment modeled by negatively charged (SDS) or zwitterionic (DPC) micelles. The data presented in Figure 3 demonstrates that all peptides adopt an α-helical conformation in complex with SDS micelles, and 6 of 7 are helical with DPC micelles. Detailed analysis of the circular dichroism spectra (Supplementary Material, Supplementary Table S6) with deconvolution algorithms allows calculating the average α-helical content of the peptides to be 69% and 70% in complex with anionic and zwitterionic micelles, respectively, while only 3% in buffer solution without lipids.

3.7 Designed peptides exhibit high antimicrobial activity

For each peptide, we determined minimal inhibitory concentration (MIC), a standard measure of antibacterial activity, at which a peptide inhibits the growth of Gram-negative and Gram-positive bacteria, using conventional microdilution assay in Mueller–Hinton broth. Most of the synthesized peptides (5 of 7) were active, and three of them had MIC values below 60 μM, which is comparable to natural AMPs (Table 1). For the most active among designed peptides, P1, MIC values are in the low micromolar range for Gram-negative pathogens *E. coli* (8 μM) and *P. aeruginosa* (2 μM), as well as Gram-positive *S. aureus* (8 μM) and *L. monocytogenes* (1 μM). This result demonstrates that the method generates new peptides with antibacterial activity similar to the most potent natural AMPs such as LL-37 and magainin 2, those MICs against *P. aeruginosa* are about 4 and 14 μM, respectively.

For the two most active candidates, P1 and P2, we synthesized and assayed four scrambled peptides lacking patterns with high

Table 1. Designed peptides exhibit high antimicrobial activity

Peptide: Sequence	E. coli ATCC 25922	P. aeruginosa ATCC 27853	S. aureus ATCC 25923	L. monocytogenes EGD
P1: **KI**GVL**KK**Y**FKI**GAL**IKA**I**IK**-NH$_2$	8	2	8	1
Scrambled P1 variants:				
KKKFIYIVLALIKGAIIKKG-NH$_2$	32	64	64	32
KGKKGVIIAILLFAIIYKKK-NH$_2$	128	128	128	128
P2: **LKK**L**K**QL**L**G**K**LSEF**AAA**FV**A**-NH$_2$	32	16	32	8
Scrambled P2 variants:				
SAKFKLAVALGQFKEKLLLA-NH$_2$		Not active (256–512)		128
KEKLAVALLFAKSKQAFGLL-NH$_2$		Not active (512–1024)		
P3: **G**Q**L**N**K**FI**KK**AQR**K**FHE**K**FA**K**-NH$_2$	128	128	128	128
P4: **K**VFKS**VV**KLLE**K**TVL**KK**F**S**K-NH$_2$	64	32	64	32
P5: **GA**LS**K**H**A**AEL**K**A**K**QRTS**L**E**K**-NH$_2$		Not active (≥256)		
P6: **LKK**LVR**K**AASIS**A**SL**AA**RH**A**-NH$_2$		Not active (≥256)		
P7: **K**A**A**KTVF**K**LF**K**LQA**K**R**A**IE**A**-NH$_2$	128	128	128	128
LL-37:				
LLGDFFRKSKEKIGKEFKRIVQRIKDFLRNLVPRTES	4	4	4	1
Magainin 2:				
GIGKFLHSAKKFGKAFVGEIMNS	0.4–2.8[a]	14–28[a]	14–28[a]	–

Notes: Experimentally determined antimicrobial activity (MIC, μM) of the designed peptides against Gram(+) and Gram(–) bacteria. For the two most active peptides, P1 and P2, we synthesized and assayed their scrambled variants, which exhibited significantly reduced activity. Data for natural AMPs LL-37 (measured in this study) and magainin 2 (from Zasloff *et al.*, 1988) is shown for comparison. Residues in bold indicate the template used to construct each peptide.

[a]Data for magainin 2 is taken from Zasloff et al. 1988.
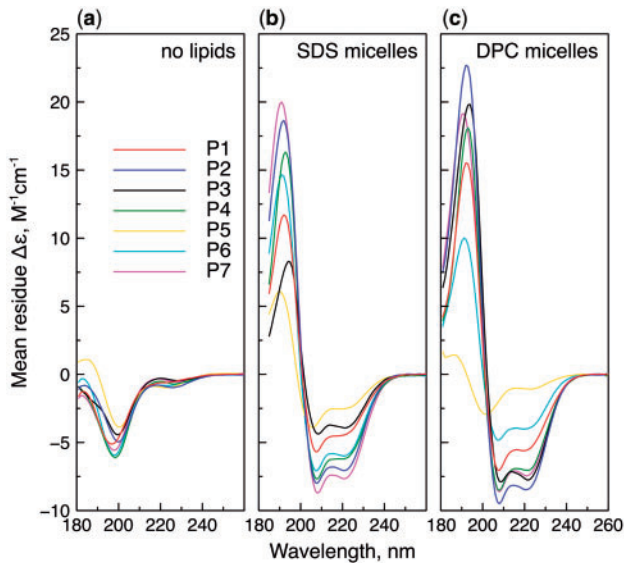
Fig. 3. Designed peptides adopt an α-helical conformation in a lipid environment. Circular dichroism spectra of the designed peptides in solution (**a**), in complex with negatively charged SDS micelles (**b**) or zwitterionic DPC micelles (**c**). All peptides except P5 were unstructured in solution but adopted an α-helical conformation in the lipid environment

support. As seen from Table 1, all scrambled variants had significantly reduced activities compared to the original peptides. This allows us to conclude that the functionality of designed peptides was determined by patterns, not solely their amino acid composition.

4 Discussion

We suggest that sequence patterns can adequately describe common structural and functional features underlying similar activities of non-homologous AMPs. Our approach may be viewed as a generalization of previous template-based methods successfully used to construct new potent α-helical peptides (Zelezetsky and Tossi, 2006) and disulfide-containing peptides (Yount and Yeaman, 2004). However, unlike these methods, our templates are designed *de novo* and expand the existing repertoire of natural peptides.

Indeed our approach is statistical like most *de novo* strategies, and some sequences generated from templates turn out non-active. For instance, previous linguistic approach yielded 18 active candidates of 40 synthetic peptides, but only 6 of them killed both Gram(+) and Gram(–) bacteria (Loose *et al.*, 2006). Aiming to understand, why some peptides constructed from patterns are highly active, while others are not, we analyzed the physicochemical properties of the designed peptides (Supplementary Material, Supplementary Table S7). We found a remarkable correlation between activity and membrane binding free energy, which is shown in Figure 4. Two non-active peptides in our set, P5 and P6, demonstrate the lowest membrane binding and lowest helicity in the membrane environment, while the most potent peptides P1 and P2 have the highest membrane binding free energies essentially identical to natural AMP magainin 2. Interestingly, other widely recognized AMP properties such as hydrophobicity, amphiphilicity and net charge virtually do not correlate with the activity of designed peptides and membrane binding free energy seem to be the principal determinant of antibacterial activity, as suggested earlier (Melo *et al.*, 2009).

The observation that designed peptides adopted helical conformation upon binding to lipid micelles allows us to suggest that patterns determine both structural class and membrane-induced folding which is a major feature of lipid–AMP interactions (Seelig, 2004; White and Wimley, 1999). This is especially important because coil to helix transition significantly contributes to membrane binding free energy. The ability to aggregate is another functional feature of AMPs, which relates them to amyloid peptides (Torrent *et al.*, 2011). We assume that some patterns may also encode aggregation-prone regions which stimulate peptide self-association on a membrane surface, ultimately leading to its disruption.
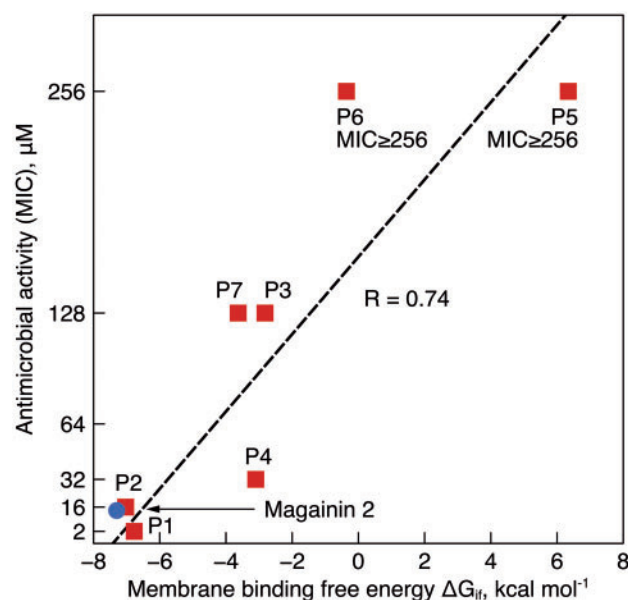
Fig. 4. Activity correlates with membrane binding free energy. MIC of the designed peptides towards *P. aeruginosa* plotted against theoretical membrane binding free energy. Linear regression and correlation coefficient (R) calculated for active peptides P1–P4 and P7. Inactive peptides P5 and P6 possess lower helicity and membrane binding free energy. Natural AMP magainin 2 (blue circle) is shown for comparison (Color version of this figure is available at *Bioinformatics* online.)

In conclusion, we anticipate that machine learning techniques, recently employed for the design of anticancer peptides (Lin *et al.*, 2015), may enforce our approach because templates significantly reduce the sequence space and large AMP databases became available as training sets. Then a template would explicitly define the structural and functional scaffold of a synthetic peptide, while machine learning would fine tune its sequence to generate *in silico* matured AMPs with enhanced activity.

References

Chen,Y. *et al.* (2005) Rational design of α-helical antimicrobial peptides with enhanced activities and specificity/therapeutic index. *J. Biol. Chem.*, **280**, 12316–12329.

Cherkasov,A. *et al.* (2009) Use of artificial intelligence in the design of small peptide antibiotics effective against a broad spectrum of highly antibiotic-resistant superbugs. *ACS Chem. Biol.*, **4**, 65–74.

Finkelstein,A.V. and Ptitsyn,O.B. (2002) *Protein Physics: A Course of Lectures*. Academic Press, New York.

Fjell,C.D. *et al.* (2011) Designing antimicrobial peptides: form follows function. *Nat. Rev. Drug Discov.*, **11**, 37–51.

Hilpert,K. *et al.* (2005) High-throughput generation of small antibacterial peptides with improved activity. *Nat. Biotechnol.*, **23**, 1008–1012.

Jones,D.T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.*, **292**, 195–202.

Kawasaki,H. and Iwamuro,S. (2008) Potential roles of histones in host defense as antimicrobial agents. *Infect. Disord. Drug Targets*, **8**, 195–205.

Li,W. and Godzik,A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**, 1658–1659.

Lin,Y.C. *et al.* (2015) Multidimensional design of anticancer peptides. *Angew. Chem. Int. Ed.*, **54**, 10370–10374.

Loose,C. *et al.* (2006) A linguistic model for the rational design of antimicrobial peptides. *Nature*, **443**, 867–869.

Melo,M.N. *et al.* (2009) Antimicrobial peptides: linking partition, activity and high membrane-bound concentrations. *Nat. Rev. Microbiol.*, **7**, 245–250.

Nguyen,L.T. *et al.* (2011) The expanding scope of antimicrobial peptide structures and their modes of action. *Trends Biotechnol.*, **29**, 464–472.

Peschel,A. and Sahl,H.G. (2006) The co-evolution of host cationic antimicrobial peptides and microbial resistance. *Nat. Rev. Microbiol.*, **4**, 529–536.

Powers,J.P.S. and Hancock,R.E.W. (2003) The relationship between peptide structure and antibacterial activity. *Peptides*, **24**, 1681–1691.

Rigoutsos,I. and Floratos,A. (1998) Combinatorial pattern discovery in biological sequences: the TEIRESIAS algorithm. *Bioinformatics*, **14**, 55–67.

Seelig,J. (2004) Thermodynamics of lipid-peptide interactions. *Biochim. Biophys. Acta*, **1666**, 40–50.

Torrent,M. *et al.* (2011) The generation of antimicrobial peptide activity: a trade-off between charge and aggregation? *Angew. Chem. Int. Ed.*, **50**, 10686–10689.

Tossi,A. *et al.* (1997) Design of synthetic antimicrobial peptides based on sequence analogy and amphipathicity. *Eur. J. Biochem.*, **250**, 549–558.

Wang,G. *et al.* (2009) APD2: the updated antimicrobial peptide database and its application in peptide design. *Nucleic Acids Res.*, **37**, D933–D937.

Wang,G. *et al.* (2016) APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Res.*, **44**, D1087–D1093.

White,S.H. and Wimley,W.C. (1999) Membrane protein folding and stability: physical principles. *Annu. Rev. Biophys. Biomol. Struct.*, **28**, 319–365.

Yount,N.Y. and Yeaman,M.R. (2004) Multidimensional signatures in antimicrobial peptides. *Proc. Natl. Acad. Sci. USA*, **101**, 7363–7368.

Zasloff,M. (2002) Antimicrobial peptides of multicellular organisms. *Nature*, **415**, 389–395.

Zasloff,M. *et al.* (1988) Antimicrobial activity of synthetic magainin peptides and several analogues. *Proc. Natl. Acad. Sci. USA*, **85**, 910–913.

Zelezetsky,I. and Tossi,A. (2006) Alpha-helical antimicrobial peptides—using a sequence template to guide structure-activity relationship studies. *Biochim. Biophys. Acta*, **1758**, 1436–1449.